INTRODUCTION TO DATA MIGRATION

Trish Rose-Sandler
VRA Conference Kansas City, MO
March 28 2007

Definition

“Data migration is the *transferring of data* between storage types, formats, or computer systems. Data migration is usually performed programmatically to achieve an *automated migration*, freeing up human resources from tedious tasks. It is required when organizations or individuals change computer systems or upgrade to new systems.”

Wikipedia 2/21/07
3 Stages of Data Migration

- **Pre Migration**
  - Analyzing, Mapping, Normalizing/Transforming, Testing, Backup

- **Migration**

- **Post Migration**
  - Quality Control, Cleanup, Update Cataloging Guidelines
Pre-Migration: Data Analysis

Evaluate data for:

- **Consistency**
- **Unnecessary redundancy across records**
- **Identify errors**
- **Re-evaluate relationships & structures**
Pre-Migration: Mapping

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>CreatorName</td>
<td>vra:agentName, dc:creator</td>
</tr>
<tr>
<td>Date</td>
<td>vra:date, dc:date</td>
</tr>
<tr>
<td>Site</td>
<td>vra:locationName, dc:coverage</td>
</tr>
<tr>
<td>General Notes</td>
<td>vra:description, dc: description</td>
</tr>
</tbody>
</table>

Example of Source to Target map
Pre-Migration: Mapping

Source Set

Examples of parsing
- Source data stores all agent names in creator field w/o a qualifying role
  - Creator
- Target data stores all agent names separate from but linked to roles
  - Agent.name
  - Agent.role

Pre-Migration: Mapping

Target Set

Mix of community standards and local elements
- Descriptive – e.g. Core 4.0
- Technical – e.g. Core 4.0, MIX
- Rights – e.g. Core 4.0, METS Rights
- Local – e.g. dateImagePurchased
Pre-Migration: Mapping

Familiarize yourself with standards’ documentation

Examples

VRA Core 4.0 standard
- Core 4.0 documentation
- Cataloging Cultural Objects (CCO) content standard

MODS standard
- MODS documentation guidelines
  [http://www.loc.gov/standards/mods/]
- DLF/Aquifer Implementation Guidelines for Shareable MODS Records

Pre-Migration: Mapping

Crosswalks may be useful, e.g. Getty crosswalk
Pre-Migration: Mapping

The goal of normalization is to transform or clean up your data values so they conform to accepted standards, are more consistent, and can be understood by any user of your images.

Can be done during
Pre Migration
Migration
Post Migration

Data to Normalize

Abbreviations

e.g. Material= ol on cv.; source = DOA; attribution=sch of

Inconsistencies

e.g. expressions of circa: Date= c1947; c. 1988; c 500 AD;
ca. 15th Century

Formatting conventions

e.g. title=[Grapevines at Mission de San Ignacio]
Pre-Migration: Mapping

Other types of normalization

Source system uses codes for values
  e.g. Name=Johnson, Ben  type=1 (1=personal)

Create indexed versions of dates
  e.g. 2/9/01, mid 15th century, October 1945
  YYYY-MM-DD, YYYY-MM, YYYY (ISO 8601 standard)

Pre-Migration: Mapping

Assess metadata granularity

Examples of narrow and broad worktypes

Narrow: City planning, urbanism, landscape design, garden design, environmental design

Broad: Architecture

Narrow: costume design, fashion design, clothing, jewelry, ornament, body decoration

Broad: Costume and Jewelry
Pre-Migration: Testing

Hand pick records to evaluate mapping decisions and to test for normalization and diacritics problems

Diacritics- best to encode in UTF-8 Unicode or use Unicode decimal or hexadecimal character references

Display

Karlstejn (Stredoceský kraj, Czech Republic)--Castle

Exported using Unicode decimal character references

Karls&#780;tejn (Str&#780;edoc&#780;eský kraj, Czech Republic)--Castle

Pre-Migration: Final Word

BACK-UP, BACK-UP, BACK-UP
Migration: A few words

Enlist the help of a programmer or database administrator

Need for assistance will depend on your source and target systems (what tools they provide for migrating data), how much data normalization you’ll need, restructuring of the data, etc.

A db administrator can help with target system setup (forms, reports, security, backup, etc)

Post Migration

Quality Control (QC)

Data Cleanup

Update Cataloging Guidelines